

SEQUENCE LISTING



(1) GENERAL INFORMATION:

- (i) APPLICANT: Benz, Christopher C.  
Scott, Gary K.  
Chang, Chuan-Hsiung
- (ii) TITLE OF INVENTION: A New ETS-Related Gene Overexpressed in Human Breast and Epithelial Cancers
- (iii) NUMBER OF SEQUENCES: 38
- (iv) CORRESPONDENCE ADDRESS:
  - (A) ADDRESSEE: Townsend and Townsend and Crew LLP
  - (B) STREET: Two Embarcadero Center, Eighth Floor
  - (C) CITY: San Francisco
  - (D) STATE: California
  - (E) COUNTRY: USA
  - (F) ZIP: 94111-3834
- (v) COMPUTER READABLE FORM:
  - (A) MEDIUM TYPE: Floppy disk
  - (B) COMPUTER: IBM PC compatible
  - (C) OPERATING SYSTEM: PC-DOS/MS-DOS
  - (D) SOFTWARE: PatentIn Release #1.0, Version #1.30
- (vi) CURRENT APPLICATION DATA:
  - (A) APPLICATION NUMBER: US 08/978,217
  - (B) FILING DATE: 25-NOV-1997
  - (C) CLASSIFICATION:
- (vii) PRIOR APPLICATION DATA:
  - (A) APPLICATION NUMBER: US 60/031,504
  - (B) FILING DATE: 27-NOV-1996
- (viii) ATTORNEY/AGENT INFORMATION:
  - (A) NAME: Hunter, Tom
  - (B) REGISTRATION NUMBER: 38,498
  - (C) REFERENCE/DOCKET NUMBER: 02307E-071110US
- (ix) TELECOMMUNICATION INFORMATION:
  - (A) TELEPHONE: (415) 576-0200
  - (B) TELEFAX: (415) 576-0300

(2) INFORMATION FOR SEQ ID NO:1:

- (i) SEQUENCE CHARACTERISTICS:
    - (A) LENGTH: 1116 base pairs
    - (B) TYPE: nucleic acid
    - (C) STRANDEDNESS: single
    - (D) TOPOLOGY: linear
  - (ii) MOLECULE TYPE: cDNA
  - (ix) FEATURE:
    - (A) NAME/KEY: CDS
    - (B) LOCATION: 1..1116
    - (D) OTHER INFORMATION: /product= "human ESX"
- /note= "epithelial-restricted with serine box (ESX)"

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

ATG	GCT	GCA	ACC	TGT	GAG	ATT	AGC	AAC	ATT	TTT	AGC	AAC	TAC	TTC	AGT	48
Met	Ala	Ala	Thr	Cys	Glu	Ile	Ser	Asn	Ile	Phe	Ser	Asn	Tyr	Phe	Ser	
1				5					10					15		
GCG	ATG	TAC	AGC	TCG	GAG	GAC	TCC	ACC	CTG	GCC	TCT	GTT	CCC	CCT	GCT	96
Ala	Met	Tyr	Ser	Ser	Glu	Asp	Ser	Thr	Leu	Ala	Ser	Val	Pro	Pro	Ala	
			20					25					30			
GCC	ACC	TTT	GGG	GCC	GAT	GAC	TTG	GTA	CTG	ACC	CTG	AGC	AAC	CCC	CAG	144
Ala	Thr	Phe	Gly	Ala	Asp	Asp	Leu	Val	Leu	Thr	Leu	Ser	Asn	Pro	Gln	
		35					40					45				
ATG	TCA	TTG	GAG	GGT	ACA	GAG	AAG	GCC	AGC	TGG	TTG	GGG	GAA	CAG	CCC	192
Met	Ser	Leu	Glu	Gly	Thr	Glu	Lys	Ala	Ser	Trp	Leu	Gly	Glu	Gln	Pro	
	50					55					60					
CAG	TTC	TGG	TCG	AAG	ACG	CAG	GTT	CTG	GAC	TGG	ATC	AGC	TAC	CAA	GTG	240
Gln	Phe	Trp	Ser	Lys	Thr	Gln	Val	Leu	Asp	Trp	Ile	Ser	Tyr	Gln	Val	
65					70					75					80	
GAG	AAG	AAC	AAG	TAC	GAC	GCA	AGC	GCC	ATT	GAC	TTC	TCA	CGA	TGT	GAC	288
Glu	Lys	Asn	Lys	Tyr	Asp	Ala	Ser	Ala	Ile	Asp	Phe	Ser	Arg	Cys	Asp	
				85					90					95		
ATG	GAT	GGC	GCC	ACC	CTC	TGC	AAT	TGT	GCC	CTT	GAG	GAG	CTG	CGT	CTG	336
Met	Asp	Gly	Ala	Thr	Leu	Cys	Asn	Cys	Ala	Leu	Glu	Glu	Leu	Arg	Leu	
			100					105					110			
GTC	TTT	GGG	CCT	CTG	GGG	GAC	CAA	CTC	CAT	GCC	CAG	CTG	CGA	GAC	CTC	384
Val	Phe	Gly	Pro	Leu	Gly	Asp	Gln	Leu	His	Ala	Gln	Leu	Arg	Asp	Leu	
		115					120					125				
ACT	TCC	AGC	TCT	TCT	GAT	GAG	CTC	AGT	TGG	ATC	ATT	GAG	CTG	CTG	GAG	432
Thr	Ser	Ser	Ser	Ser	Asp	Glu	Leu	Ser	Trp	Ile	Ile	Glu	Leu	Leu	Glu	
	130					135					140					
AAG	GAT	GGC	ATG	GCC	TTC	CAG	GAG	GCC	CTA	GAC	CCA	GGG	CCC	TTT	GAC	480
Lys	Asp	Gly	Met	Ala	Phe	Gln	Glu	Ala	Leu	Asp	Pro	Gly	Pro	Phe	Asp	
145					150					155					160	
CAG	GGC	AGC	CCC	TTT	GCC	CAG	GAG	CTG	CTG	GAC	GAC	GGT	CAG	CAA	GCC	528
Gln	Gly	Ser	Pro	Phe	Ala	Gln	Glu	Leu	Leu	Asp	Asp	Gly	Gln	Gln	Ala	
				165					170					175		
AGC	CCC	TAC	CAC	CCC	GGC	AGC	TGT	GGC	GCA	GGA	GCC	CCC	TCC	CCT	GGC	576
Ser	Pro	Tyr	His	Pro	Gly	Ser	Cys	Gly	Ala	Gly	Ala	Pro	Ser	Pro	Gly	
			180					185					190			
AGC	TCT	GAC	GTC	TCC	ACC	GCA	GGG	ACT	GGT	GCT	TCT	CGG	AGC	TCC	CAC	624
Ser	Ser	Asp	Val	Ser	Thr	Ala	Gly	Thr	Gly	Ala	Ser	Arg	Ser	Ser	His	
		195					200					205				
TCC	TCA	GAC	TCC	GGT	GGA	AGT	GAC	GTG	GAC	CTG	GAT	CCC	ACT	GAT	GGC	672
Ser	Ser	Asp	Ser	Gly	Gly	Ser	Asp	Val	Asp	Leu	Asp	Pro	Thr	Asp	Gly	
	210					215					220					
AAG	CTC	TTC	CCC	AGC	GAT	GGT	TTT	CGT	GAC	TGC	AAG	AAG	GGG	GAT	CCC	720
Lys	Leu	Phe	Pro	Ser	Asp	Gly	Phe	Arg	Asp	Cys	Lys	Lys	Gly	Asp	Pro	
225					230					235					240	
AAG	CAC	GGG	AAG	CGG	AAA	CGA	GGC	CGG	CCC	CGA	AAG	CTG	AGC	AAA	GAG	768
Lys	His	Gly	Lys	Arg	Lys	Arg	Gly	Arg	Pro	Arg	Lys	Leu	Ser	Lys	Glu	
				245					250					255		

TAC	TGG	GAC	TGT	CTC	GAG	GGC	AAG	AAG	AGC	AAG	CAC	GCG	CCC	AGA	GGC	816
Tyr	Trp	Asp	Cys	Leu	Glu	Gly	Lys	Lys	Ser	Lys	His	Ala	Pro	Arg	Gly	
			260					265					270			
ACC	CAC	CTG	TGG	GAG	TTC	ATC	CGG	GAC	ATC	CTC	ATC	CAC	CCG	GAG	CTC	864
Thr	His	Leu	Trp	Glu	Phe	Ile	Arg	Asp	Ile	Leu	Ile	His	Pro	Glu	Leu	
		275					280					285				
AAC	GAG	GGC	CTC	ATG	AAG	TGG	GAG	AAT	CGG	CAT	GAA	GGC	GTC	TTC	AAG	912
Asn	Glu	Gly	Leu	Met	Lys	Trp	Glu	Asn	Arg	His	Glu	Gly	Val	Phe	Lys	
	290					295					300					
TTC	CTG	CGC	TCC	GAG	GCT	GTG	GCC	CAA	CTA	TGG	GGC	CAA	AAG	AAA	AAG	960
Phe	Leu	Arg	Ser	Glu	Ala	Val	Ala	Gln	Leu	Trp	Gly	Gln	Lys	Lys	Lys	
305				310						315					320	
AAC	AGC	AAC	ATG	ACC	TAC	GAG	AAG	CTG	AGC	CGG	GCC	ATG	AGG	TAC	TAC	1008
Asn	Ser	Asn	Met	Thr	Tyr	Glu	Lys	Leu	Ser	Arg	Ala	Met	Arg	Tyr	Tyr	
			325						330					335		
TAC	AAA	CGG	GAG	ATC	CTG	GAA	CGG	GTG	GAT	GGC	CGG	CGA	CTC	GTC	TAC	1056
Tyr	Lys	Arg	Glu	Ile	Leu	Glu	Arg	Val	Asp	Gly	Arg	Arg	Leu	Val	Tyr	
			340					345					350			
AAG	TTT	GGC	AAA	AAC	TCA	AGC	GGC	TGG	AAG	GAG	GAA	GAG	GTT	CTC	CAG	1104
Lys	Phe	Gly	Lys	Asn	Ser	Ser	Gly	Trp	Lys	Glu	Glu	Glu	Val	Leu	Gln	
		355					360					365				
AGT	CGG	AAC	TGA													1116
Ser	Arg	Asn														
		370														

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 371 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

Met	Ala	Ala	Thr	Cys	Glu	Ile	Ser	Asn	Ile	Phe	Ser	Asn	Tyr	Phe	Ser	
1				5					10					15		
Ala	Met	Tyr	Ser	Ser	Glu	Asp	Ser	Thr	Leu	Ala	Ser	Val	Pro	Pro	Ala	
			20					25					30			
Ala	Thr	Phe	Gly	Ala	Asp	Asp	Leu	Val	Leu	Thr	Leu	Ser	Asn	Pro	Gln	
		35					40					45				
Met	Ser	Leu	Glu	Gly	Thr	Glu	Lys	Ala	Ser	Trp	Leu	Gly	Glu	Gln	Pro	
	50					55					60					
Gln	Phe	Trp	Ser	Lys	Thr	Gln	Val	Leu	Asp	Trp	Ile	Ser	Tyr	Gln	Val	
65					70					75					80	
Glu	Lys	Asn	Lys	Tyr	Asp	Ala	Ser	Ala	Ile	Asp	Phe	Ser	Arg	Cys	Asp	
				85					90					95		
Met	Asp	Gly	Ala	Thr	Leu	Cys	Asn	Cys	Ala	Leu	Glu	Glu	Leu	Arg	Leu	
			100					105					110			
Val	Phe	Gly	Pro	Leu	Gly	Asp	Gln	Leu	His	Ala	Gln	Leu	Arg	Asp	Leu	
		115					120					125				

Thr Ser Ser Ser Ser Asp Glu Leu Ser Trp Ile Ile Glu Leu Leu Glu  
 130 135 140  
 Lys Asp Gly Met Ala Phe Gln Glu Ala Leu Asp Pro Gly Pro Phe Asp  
 145 150 155 160  
 Gln Gly Ser Pro Phe Ala Gln Glu Leu Leu Asp Asp Gly Gln Gln Ala  
 165 170 175  
 Ser Pro Tyr His Pro Gly Ser Cys Gly Ala Gly Ala Pro Ser Pro Gly  
 180 185 190  
 Ser Ser Asp Val Ser Thr Ala Gly Thr Gly Ala Ser Arg Ser Ser His  
 195 200 205  
 Ser Ser Asp Ser Gly Gly Ser Asp Val Asp Leu Asp Pro Thr Asp Gly  
 210 215 220  
 Lys Leu Phe Pro Ser Asp Gly Phe Arg Asp Cys Lys Lys Gly Asp Pro  
 225 230 235 240  
 Lys His Gly Lys Arg Lys Arg Gly Arg Pro Arg Lys Leu Ser Lys Glu  
 245 250 255  
 Tyr Trp Asp Cys Leu Glu Gly Lys Lys Ser Lys His Ala Pro Arg Gly  
 260 265 270  
 Thr His Leu Trp Glu Phe Ile Arg Asp Ile Leu Ile His Pro Glu Leu  
 275 280 285  
 Asn Glu Gly Leu Met Lys Trp Glu Asn Arg His Glu Gly Val Phe Lys  
 290 295 300  
 Phe Leu Arg Ser Glu Ala Val Ala Gln Leu Trp Gly Gln Lys Lys Lys  
 305 310 315 320  
 Asn Ser Asn Met Thr Tyr Glu Lys Leu Ser Arg Ala Met Arg Tyr Tyr  
 325 330 335  
 Tyr Lys Arg Glu Ile Leu Glu Arg Val Asp Gly Arg Arg Leu Val Tyr  
 340 345 350  
 Lys Phe Gly Lys Asn Ser Ser Gly Trp Lys Glu Glu Glu Val Leu Gln  
 355 360 365  
 Ser Arg Asn  
 370

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1907 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:

- (A) NAME/KEY: CDS
  - (B) LOCATION: 96..1211
  - (D) OTHER INFORMATION: /product= "human ESX"
- /note= "epithelial-restricted with serine box (ESX)"

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

CGGCCAGATA CCTCAGCGCT ACCTGGCGGA ACTGGATTTC TCTCCCGCCT GCCGGCCTGC	60
CTGCCACAGC CGGACTCCGC CACTCCGGTA GCCTCATGGC TGCAACCTGT GAGATTAGCA	120
ACATTTTTAG CAACTACTTC AGTGCGATGT ACAGCTCGGA GGACTCCACC CTGGCCTCTG	180
TTCCCCCTGC TGCCACCTTT GGGGCCGATG ACTTGGTACT GACCCTGAGC AACCCCCAGA	240
TGTCATTGGA GGGTACAGAG AAGGCCAGCT GGTGGGGGA ACAGCCCCAG TTCTGGTCGA	300
AGACGCAGGT TCTGGACTGG ATCAGCTACC AAGTGGAGAA GAACAAGTAC GACGCAAGCG	360
CCATTGACTT CTCACGATGT GACATGGATG GCGCCACCCT CTGCAATTGT GCCCTTGAGG	420
AGCTGCGTCT GGTCTTTGGG CCTCTGGGGG ACCAACTCCA TGCCCAGCTG CGAGACCTCA	480
CTTCCAGCTC TTCTGATGAG CTCAGTTGGA TCATTGAGCT GCTGGAGAAG GATGGCATGG	540
CCTTCCAGGA GGCCCTAGAC CCAGGGCCCT TTGACCAGGG CAGCCCCTTT GCCCAGGAGC	600
TGCTGGACGA CGGTACAGAA GCCAGCCCCCT ACCACCCCGG CAGCTGTGGC GCAGGAGCCC	660
CCTCCCCTGG CAGCTCTGAC GTCTCCACCG CAGGGACTGG TGCTTCTCGG AGCTCCCCT	720
CCTCAGACTC CGGTGGAAGT GACGTGGACC TGGATCCCAC TGATGGCAAG CTCTTCCCCA	780
GCGATGGTTT TCGTGA CTGC AAGAAGGGG ATCCCAAGCA CGGGAAGCGG AAACGAGGCC	840
GGCCCCGAAA GCTGAGCAA GAGTACTGGG ACTGTCTCGA GGGCAAGAAG AGCAAGCACG	900
CGCCCAGAGG CACCCACCTG TGGGAGTTCA TCCGGGACAT CCTCATCCAC CCGGAGCTCA	960
ACGAGGGCCT CATGAAGTGG GAGAATCGGC ATGAAGGCGT CTTCAAGTTC CTGCGCTCCG	1020
AGGCTGTGGC CCAACTATGG GGCCAAAAGA AAAAGAACAG CAACATGACC TACGAGAAGC	1080
TGAGCCGGGC CATGAGGTAC TACTACAAAC GGGAGATCCT GGAACGGGTG GATGGCCGGC	1140
GACTCGTCTA CAAGTTTGGC AAAA ACTCAA GCGGCTGGAA GGAGGAAGAG GTTCTCCAGA	1200
GTCGGA ACTG AGGGTTGGAA CTATACCCGG GACC AA CTG ACGGACCACT CGAGGCCTGC	1260
AAACCTTCCT GGGAGGACAG GCAGGCCAGA TGGCCCCCTCC ACTGGGGAAT GCTCCCAGCT	1320
GTGCTGTGGA GAGAAGCTGA TGTTTTGGTG TATTGTCAGC CATCGTCCTT GGACTCGGAG	1380
ACTATGGCCT CGCCTCCCCA CCCTCCTCTT GGAATTACAA GCCCTGGGGT TTGAAGCTGA	1440
CTTTATAGCT GCAAGTGTAT CTCCTTTTAT CTGGTGCCTC CTCAAACCCA GTCTCAGACA	1500
CTTAAATGCA GACAACACCT TCTTCCTGCA GACACTTGGA CTGAGCCAAG GAGGCTTGGG	1560
AGGCCCTAGG GAGCACCGTG ATGGAGAGGA CAGAGCAGGG GCTCCAGCAC TTCTTTCTGG	1620
ACTGGCGTTC ACCTCCCTGC TCAGTGCTTG GGCTCCACGG GCAGGGGTCA GAGCACTCCC	1680
TAATTTATGT GCTATATAAA TATGTCAGAT GTACATAGAG ATCTATTTTT TCTAAAACAT	1740
TCCCCTCCCC ACTCCTCTCC CACAGAGTGC TGGACTGTTC CAGGCCCTCC AGTGGGCTGA	1800
TGCTGGGACC CTTAGGATGG GGCTCCCAGC TCCTTTCTCC TGTGAATGGA GGCAGAGACC	1860
TCCAATAAAG TGCCTTCTGG GCTTTTTCTA AAAAAAAAAA AAAAAA	1907

(2) INFORMATION FOR SEQ ID NO:4:

- (i) SEQUENCE CHARACTERISTICS:  
    (A) LENGTH: 189 base pairs  
    (B) TYPE: nucleic acid  
    (C) STRANDEDNESS: single  
    (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

- (ix) FEATURE:  
    (A) NAME/KEY: -  
    (B) LOCATION: 1..189  
    (D) OTHER INFORMATION: /note= "first variable region  
(nucleotides 1-189 of SEQ ID NO:1)"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

```
ATGGCTGCAA CCTGTGAGAT TAGCAACATT TTTAGCAACT ACTTCAGTGC GATGTACAGC      60
TCGGAGGACT CCACCCTGGC CTCTGTTCCC CCTGCTGCCA CCTTTGGGGC CGATGACTTG      120
GTACTGACCC TGAGCAACCC CCAGATGTCA TTGGAGGGTA CAGAGAAGGC CAGCTGGTTG      180
GGGGAACAG                                     189
```

(2) INFORMATION FOR SEQ ID NO:5:

- (i) SEQUENCE CHARACTERISTICS:  
    (A) LENGTH: 120 base pairs  
    (B) TYPE: nucleic acid  
    (C) STRANDEDNESS: single  
    (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

- (ix) FEATURE:  
    (A) NAME/KEY: -  
    (B) LOCATION: 1..120  
    (D) OTHER INFORMATION: /note= "pointed region (nucleotides  
190-309 of SEQ ID NO:1)"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

```
CCCCAGTTCT GGTCGAAGAC GCAGGTTCTG GACTGGATCA GCTACCAAGT GGAGAAGAAC      60
AAGTACGACG CAAGCGCCAT TGA CTCTCTCA CGATGTGACA TGGATGGCGC CACCCTCTGC      120
```

(2) INFORMATION FOR SEQ ID NO:6:

- (i) SEQUENCE CHARACTERISTICS:  
    (A) LENGTH: 252 base pairs  
    (B) TYPE: nucleic acid  
    (C) STRANDEDNESS: single  
    (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(ix) FEATURE:  
 (A) NAME/KEY: -  
 (B) LOCATION: 1..252  
 (D) OTHER INFORMATION: /note= "second variable region  
 (nucleotides 310-561 of SEQ ID NO:1)"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

AATTGTGCCC TTGAGGAGCT GCGTCTGGTC TTTGGGCTC TGGGGGACCA ACTCCATGCC	60
CAGCTGCGAG ACCTCACTTC CAGCTCTTCT GATGAGCTCA GTTGGATCAT TGAGCTGCTG	120
GAGAAGGATG GCATGGCCTT CCAGGAGGCC CTAGACCCAG GGCCCTTTGA CCAGGGCAGC	180
CCCTTTGCCC AGGAGCTGCT GGACGACGGT CAGCAAGCCA GCCCCTACCA CCCCGGCAGC	240
TGTGGCGCAG GA	252

(2) INFORMATION FOR SEQ ID NO:7:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 84 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS:  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(ix) FEATURE:  
 (A) NAME/KEY: Peptide  
 (B) LOCATION: 1..84  
 (D) OTHER INFORMATION: /note= "second variable region  
 (amino acids 104-187 of SEQ ID NO:2)"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

Asn Cys Ala Leu Glu Glu Leu Arg Leu Val Phe Gly Pro Leu Gly Asp	
1 5 10 15	
Gln Leu His Ala Gln Leu Arg Asp Leu Thr Ser Ser Ser Ser Asp Glu	
20 25 30	
Leu Ser Trp Ile Ile Glu Leu Leu Glu Lys Asp Gly Met Ala Phe Gln	
35 40 45	
Glu Ala Leu Asp Pro Gly Pro Phe Asp Gln Gly Ser Pro Phe Ala Gln	
50 55 60	
Glu Leu Leu Asp Asp Gly Gln Gln Ala Ser Pro Tyr His Pro Gly Ser	
65 70 75 80	
Cys Gly Ala Gly	

(2) INFORMATION FOR SEQ ID NO:8:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 153 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(ix) FEATURE:  
 (A) NAME/KEY: -  
 (B) LOCATION: 1..153  
 (D) OTHER INFORMATION: /note= "serine-rich region  
 (nucleotides 562-714 of SEQ ID NO:1)"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

GCCCCCTCCC CTGGCAGCTC TGACGTCTCC ACCGCAGGGA CTGGTGCTTC TCGGAGCTCC	60
CACTCCTCAG ACTCCGGTGG AAGTGACGTG GACCTGGATC CCACTGATGG CAAGCTCTTC	120
CCCAGCGATG GTTTTCGTGA CTGCAAGAAG GGG	153

(2) INFORMATION FOR SEQ ID NO:9:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 105 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(ix) FEATURE:  
 (A) NAME/KEY: -  
 (B) LOCATION: 1..105  
 (D) OTHER INFORMATION: /note= "third variable region  
 (nucleotides 715-819 of SEQ ID NO:1)"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

GATCCCAAGC ACGGGAAGCG GAAACGAGGC CGGCCCCGAA AGCTGAGCAA AGAGTACTGG	60
GACTGTCTCG AGGGCAAGAA GAGCAAGCAC GCGCCCAGAG GCACC	105

(2) INFORMATION FOR SEQ ID NO:10:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 243 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(ix) FEATURE:  
 (A) NAME/KEY: -  
 (B) LOCATION: 1..243  
 (D) OTHER INFORMATION: /note= "Ets DNA binding domain  
 (nucleotides 820-1062 of SEQ ID NO:1)"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

CACCTGTGGG AGTTCATCCG GGACATCCTC ATCCACCCGG AGCTCAACGA GGGCCTCATG	60
AAGTGGGAGA ATCGGCATGA AGGCGTCTTC AAGTTCCTGC GCTCCGAGGC TGTGGCCCAA	120
CTATGGGGCC AAAAGAAAAA GAACAGCAAC ATGACCTACG AGAAGCTGAG CCGGGCCATG	180



AGGTACTACT ACAAAACGGGA GATCCTGGAA CGGGTGGATG GCCGGCGACT CGTCTACAAG 240  
 TTT 243

(2) INFORMATION FOR SEQ ID NO:11:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 51 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(ix) FEATURE:

- (A) NAME/KEY: -
- (B) LOCATION: 1..51
- (D) OTHER INFORMATION: /note= "fourth variable region (nucleotides 1063-1113 of SEQ ID NO:1)"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

GGCAAAACT CAAGCGGCTG GAAGGAGGAA GAGGTTCTCC AGAGTCGGAA C 51

(2) INFORMATION FOR SEQ ID NO:12:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 16 amino acids
  - (B) TYPE: amino acid
  - (C) STRANDEDNESS:
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(ix) FEATURE:

- (A) NAME/KEY: Peptide
- (B) LOCATION: 1..16
- (D) OTHER INFORMATION: /note= "C-terminal 16 amino acids (amino acids 356-371 or SEQ ID NO:2)"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

Lys	Asn	Ser	Ser	Gly	Trp	Lys	Glu	Glu	Glu	Val	Leu	Gln	Ser	Arg	Asn
1				5					10					15	

(2) INFORMATION FOR SEQ ID NO:13:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 21 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(ix) FEATURE:

- (A) NAME/KEY: -
- (B) LOCATION: 1..21
- (D) OTHER INFORMATION: /note= "5' ESX-DBD primer"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

CCGGGACATC CTCATCCACC C

21

(2) INFORMATION FOR SEQ ID NO:14:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 21 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(ix) FEATURE:

- (A) NAME/KEY: -  
 (B) LOCATION: 1..21  
 (D) OTHER INFORMATION: /note= "3' ESX-DBD primer"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

GTACCTCATG GCCCGGCTCA G

21

(2) INFORMATION FOR SEQ ID NO:15:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 7752 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:

- (A) NAME/KEY: CDS  
 (B) LOCATION: join(3604..3763, 4152..4373, 4504..4599, 4788  
 ..4907, 5055..5144, 5287..5403, 6257..6452, 7001  
 ..7112)  
 (D) OTHER INFORMATION: /product= "mouse ESX"  
/note= "mouse epithelial-restricted with serine box (ESX)"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

GGATCCTTCC AAGGCACTGA CCTCACCCAA TTCTTTCTCA CTTTCTCCT CCATTTAACT	60
GTGGACGGAA TCAATACTCA GGGGGATGCG CTAGCTCTAA GATTTCTGCA GCTTTGCCTC	120
TCCTGAGCGG AAGCCCCGTG AAGGCAAGGG AGCTAGCTGA TGGACTCTTT GTGGTCTTCT	180
TCCTCTTTGC TCTGGAGACC CAACCAGGTG TTCTTAGGGG AAGGAGCACG TGAGTAGCCA	240
AGAGGCTAAA AGCTGGTTCT CCCACATTCC AGGGTAAGTG ACTGGGTAGA GGGTGTGTCT	300
GCCTCAGGCT GCTTGAGGA GGTCCCCTGA AGGGCCATGA GAAAATCCTA CCCAGAGCCC	360
TTGGTTTTCC AGCAGCCCTC CACCTAGAGG AAAGGAGCCT GTCGTTCTGA AGATGAAGAG	420
TGGAGCCTAT GGGGGTGGGC AGATTGTGTC CTGGGACAAT GGGGTACCTA GAAGAGAAAG	480
GAATCTCCTT TCGTTTGAGG TCTACCTGGG GGTCTGTGT CTGTAAATGG GGTGGAGAGA	540

GGAGAAGACA	CAGATCTTAT	AACGTAGATG	CAGGAAATGC	TGACAGTTCA	GTGTAGAGAA	600
CTTACTCAAT	TCATATAGCC	TCCAAAGCTA	TCTCCTCAGG	CAACGCAAAA	CAAACCAGTT	660
GGAGCCGCAA	GACATCTAAT	GGCTTATCGA	GTCCCACACC	CTCGATTCTT	TGCTAATTTT	720
ATGGT'TTTGC	TTTTGAGACA	ATCTACTGTA	GCCTAAGATA	GCCCCAAACT	CAAATGTAGC	780
TGAGGCTGAC	TGACCCTGAG	CTCTGGAATT	CCAGACACAT	GCATATCTTT	TGCTAGGCAA	840
TAATCGCTCT	ACCAGCTGTA	CTCCCACATT	CCAGGGTAAG	TGACTGGAAT	TCTCACTTAC	900
TATATCCCTT	TAAAAATTCC	CTGAGTGGA	TGGTTGTAGC	CAGAGGGAAA	AGGCACCAAC	960
AACTGCTTGT	CACTTTCCAA	ATTTGGTAGC	CTGAACAAAC	CACTTATCAA	GACAACAAC	1020
ATATATCATT	TCTTTTCTTC	TCTCTCTCTC	TCTCTCTCTC	TCTCTCTCTC	TCTCTCTCTC	1080
TCTCTCTCTC	TCTCTCTTTN	GAAAGAGTCT	CACTACTATG	TAGCCCTTGA	TAACCTAGAA	1140
CTCACTATGT	AGTCCAGGCT	TGGCCTTCAG	CTCGCAGAGG	TCCACTTGCC	TTGGGAGTTG	1200
AGAGATTAAA	GGGATGCATC	TCCACATGTG	TCCAACAGTG	CTTTTAAAA	ATATTTTAA	1260
AACCATGCTT	ACAGCCAGGC	ATAGTGGGCG	TGCCTTTAAT	CCCAGTACTG	GGGAGGCAGA	1320
GGTAGGTAGA	GTTCTGAGTT	GGAGGCTAGC	CACATAGTAA	GTCCCAGGAT	AGCTAGAACT	1380
ATGTAAAGAC	CATGTCTCAA	AAAAGATGCA	CACACACATA	TACACACACA	CGTTTGTATG	1440
TGTTTGT'TTA	GTGTGTATGT	GTGTGTACAC	TTGCACATAA	AGGTCAGAGT	ACCACATTAC	1500
AGGAGTCAGT	TTTCTCCTTT	TATCATGTAT	GGATGGAACA	CGGGTCCATC	CATAGCATCC	1560
TTAGCAGCAG	GTATCCTTAT	CCACTGAGCT	ATCTCAGCAG	CCCCACATTG	CTTATTGGAT	1620
G'TTTTTGGAT	GAGGATAGTT	ATATTAAAAA	GGTTTCTGGT	GTTGGTCTGG	GTAGTTACCC	1680
TTTAACCCAT	CTCTAGAGCC	TGTCTCTTGA	GTTTGAGGCC	AGCCTGGTAT	ATGTAGCTAG	1740
ACAAAGTTTC	AAAAATGAAC	AGAATCCTGG	GACTAGAACC	CATTTGTAGA	ATGCTTGCAT	1800
AAGAAGCTCT	GGGTTCAACT	TCCTGCATCT	CCAGAGGGAT	TTTGTTCTGT	AGTTTTAGTT	1860
TTTCAAGACA	GAGTTTCTCT	GTGTAGCCCT	GGCTGTCTCG	GAACCTCACTC	TGTAGACAAG	1920
GCTGGCCTCG	AACTCAGAAA	TCCTTCTACC	TCTACTTCAG	GAGTGCTGGG	ATTAAAGATG	1980
TGCGCTGCCC	TCCTCCACCC	CAATTTGTTT	TTGTTTTTTA	AGGGCCCCGG	TAAACAGTAA	2040
ATTAACATGT	GCATCCTGTT	TGTCTTTGTA	ATGACTCAAA	TGTTGGGCTT	CTGACCACTA	2100
GAGGGCAGCA	GGCAGATACT	AATGGACTGG	GCGGAGAGAA	GGGTAATCAG	GAGCAGACCA	2160
GACTCGCGGA	TAAACCAAAC	AGCACCGCCA	GCCGACCCTA	GGCGAGGAGA	GCGCCACAGG	2220
CACCAAGGGA	AGACTTGAAG	TAGTGTCTGA	TCTCTACCGC	TTCAGCAACC	ATCGCGTTTG	2280
GGTGGGCTCC	AGACAGGCAA	AGTGCCAGCA	AATGGTCCCT	GTAGCTGACT	AAACAGACTA	2340
TCAGACCCAA	ACCACCACTG	GACCGTGAAT	GTTGCCCAGT	GTGTTGCCTA	GCCGCTTTCA	2400
GAATCCAGC	TTCTGGGTGT	TGTGGAGGAA	ACCCCTTAGC	CTCGGTAAC	TTCACCAGGC	2460
CCTTCTTGTC	TCTAGACATC	TAGACAGTTG	GAAGCATCAG	TCTTGACCCA	GCCACCGGTT	2520
CAGATTCTTT	GCCTTGCTTT	TTCTTCCCCA	GTTCAGCCCT	GGCCAGGCC	CCAGGAAGAA	2580

TTTCCAGGGC CAGAGGGCAG CCTAAGGCAC AGATGCCCAC CCCTGCAATG TTCCCGCCAC	2640
ATGCCCAGTT CAGTACCCAG GGCCCAACCC CAGAGGGTGC GGAATGACAG ATTCTGACAA	2700
TCATTAAACC AGCCAGGCCT GATTTCCCAG CACCGCCCGT TAGGATATGG GCCAAGTGGC	2760
ACGGAATATG CAAATCACAT GGGACAGGGA GCCCAGTCTG AAGGCCAGGA AATCCCCAGC	2820
ATCCAATGAG CCACCAGCTC AGGTTACAAC CGGGGACGTA CGCCGAAGAC CTGGAGGGGA	2880
GGAGCTCCTG CTTTGCTCTA TTTAGAGCGG GTGGGGGCAG CGCCCTGGCC AACTCATCA	2940
CTGCTACCTG CGGAGCCTTC GACCGCTTAG ATTTTTTCCC TTCCTGTGGC CTCAGAAGCC	3000
TGCTCACCCG CCTGCCACAC CGAACCTGA CACACCTCGG TACGGTCACA TTCCCTAACT	3060
CTGGCTCCAG GAACCGTCCA GTGGATTAC AGTTCGTAAC TTAATCACTC AGGCTTGAG	3120
GTTCTAGCT GGAGTGTGG GGCTACTGTG GGTGTATTCT GGGACTGGTC AGAGACCAGA	3180
TCGGTGTCTT GGAGGGACAG GGTGGCTTCT TTGGTTCAGG AGCCACGTG ATTTGTGGAG	3240
AGACCCAGA AGAATTTGTA TCATGCTCCC ACCCGCTTGG AGATTATTTT TATTTTTCGG	3300
AGCCGAATTT CCCAGTTTGG CGCCAGCTGG CCTGACCCTT CTAGGCTCAA GAGAGCATCC	3360
AACCTCAGCT TCCCCAAGTA GCTGGCTCTT GGTGGTGATG GTGGTGGTGG TGGTGGTGGT	3420
GGTGGTGGTG GTGTGTGTGT TTGTGTGTGT GTGGTGGTGG GGGGGGTGTT GAAGAGAGAA	3480
TGTCTACAGC AACACTGAAC TTCCTGCCTC TCGGCTGTTG CTGCCCAGGC TTTGCCAGAC	3540
AGAAATGGAA GTGTATCCTG ACCTGTACCC TCCCCACCTT GTCTCCTCTT CCCAGGGGCC	3600
CTC ATG GCT GCC ACC TGT GAG ATC AGC AAC GTT TTT AGT AAC TAC TTC	3648
Met Ala Ala Thr Cys Glu Ile Ser Asn Val Phe Ser Asn Tyr Phe	
1 5 10 15	
AAC GCC ATG TAC AGC TCA GAA GAC CCC ACC CTG GCT CCT GCT CCT CCG	3696
Asn Ala Met Tyr Ser Ser Glu Asp Pro Thr Leu Ala Pro Ala Pro Pro	
20 25 30	
ACT ACC TTT GGC ACT GAA GAC TTG GTG TTG ACC CTG AAC AAC CAA CAG	3744
Thr Thr Phe Gly Thr Glu Asp Leu Val Leu Thr Leu Asn Asn Gln Gln	
35 40 45	
ATG ACA CTG GAA GGT CCA G GTGAGTGCTG TGTAAAATCT TTTCAGACAG	3793
Met Thr Leu Glu Gly Pro	
50	
GACACCAATG ATCTGAGAGG CTCTTAGATG ATAAATGGAC AGGGAGGAAG GGTATCCTGG	3853
AGTTAGTGGC TGGGGAGGAT TTATTCATTC ATATGTTTGT GTAGTACTGG GGAAAGAACC	3913
CAAACAAGAC CTTATTTATG CTAGACTGTG TTCCTAGTCC CGAGAAGACT GTACTGGCTG	3973
AGGTGGTGGG AATATAAGAA CTGTGGTGAC AGATTAAGGG AGGATGAACT TGAGAAGTAG	4033
CCATGTTGTG ATTGTGGATA TGTATCTGTC CCTCTCCGCC CCTCCTCGGG TTGTGTAGGA	4093
CCTCAGACAA GATCCCAAAG GGACAGGACT GATCCTCTGG CTGTACTCCA CCTTGCAG AG	4153
Glu	
AAG GCA AGC TGG ACT AGC GAG CGG CCC CAG TTC TGG TCG AAG ACC CAG	4201
Lys Ala Ser Trp Thr Ser Glu Arg Pro Gln Phe Trp Ser Lys Thr Gln	
55 60 65 70	

GTT CTG GAG TGG ATC AGC TAC CAA GTG GAG AAG AAC AAG TAT GAC GCC Val Leu Glu Trp Ile Ser Tyr Gln Val Glu Lys Asn Lys Tyr Asp Ala 75 80 85	4249
AGC TCC ATC GAC TTC TCC CGC TGC AAC ATG GAC GGA GCC ACC CTC TGC Ser Ser Ile Asp Phe Ser Arg Cys Asn Met Asp Gly Ala Thr Leu Cys 90 95 100	4297
AGC TGT GCG CTG GAG GAG CTG CGG CTA GTC TTT GGA CCT CTG GGA GAC Ser Cys Ala Leu Glu Glu Leu Arg Leu Val Phe Gly Pro Leu Gly Asp 105 110 115	4345
CAG CTC CAT GCC CAG CTT CGG GAC CTC A GTAAGTCTAG GCTGGGAGCC Gln Leu His Ala Gln Leu Arg Asp Leu 120 125	4393
ACAGGGCCTA AAGAGTGAGC GAGGTGGCTG GGAAGTGGGC AGGAGGGTGC AGCCATCGAG	4453
CCCCTGCCGG AACCATGGTC GGTGACGCTC TCCCTCCCTG CCTCCGCCAG CC TCC Thr Ser	4508
AAC TCT TCT GAT GAA CTC AGC TGG ATC ATC GAG CTG CTG GAG AAG GAT Asn Ser Ser Asp Glu Leu Ser Trp Ile Ile Glu Leu Leu Glu Lys Asp 130 135 140 145	4556
GGC ATG TCC TTC CAA GAG AGC CTA GGC GAC TTG GGC CCC TTT G Gly Met Ser Phe Gln Glu Ser Leu Gly Asp Leu Gly Pro Phe 150 155	4599
GTGAGAACCC ATTTTCTCCC TTTTTCCTCC CTAGCTTGTC TTGTCCCATC TGTAACCTCT	4659
CCAGAGTGCT ACAGATATTC TCTCCCAACT TGAAAATAAG TCCATAGTCA TTTCTGTGGT	4719
CCCTGGAGGG TCGTGCCTGT CCTTGCTGGT ATCCTGGGCC TCTCTAAGCT CTTAACTTCT	4779
TTTCTCAG AT CAG GGA AGT CCT TTT GCC CAG GAA CTC CTG GAT GAT GGC Asp Gln Gly Ser Pro Phe Ala Gln Glu Leu Leu Asp Asp Gly 160 165 170	4828
CGC CAG GCC AGT CCC TAC TAC TGC AGT ACC TAT GGC CCT GGA GCG CCC Arg Gln Ala Ser Pro Tyr Tyr Cys Ser Thr Tyr Gly Pro Gly Ala Pro 175 180 185	4876
TCC CCC GGC AGC TCT GAT GTC TCC ACT GCA A GTAAGTCCTG CCCTTGCCAC Ser Pro Gly Ser Ser Asp Val Ser Thr Ala 190 195	4927
AGCCTGCCTT CTCCAAGTGC CCTAGAGTGC ATCGAGTTCT TACAATACTC ATTCAGTATC	4987
TGAAGTCTGG GTACGCAGTG ACTGGGTAGG CTGGCCCTGG CATTCAAGTG GTATTCTTCA	5047
CCCCTAG GG ACC GCT ACT CCC CAG AGT TCC CAT GCC TCT GAC TCC GGT Arg Thr Ala Thr Pro Gln Ser Ser His Ala Ser Asp Ser Gly 200 205 210	5095
GGA AGT GAT GTG GAC CTG GAC CTC ACC GAG AGC AAG GTC TTC CCT AGA G Gly Ser Asp Val Asp Leu Asp Leu Thr Glu Ser Lys Val Phe Pro Arg 215 220 225	5144
GTGAGTTGAG GGCTGTTCTT GGGGGTCCTG TCCATGGGGT CTAGCCACTC CCCTCTGCCC	5204
TATGGCTGCA GTTTCTGTAC CAAGGCTCCC TGTTGACACC CTGCCCTTAC CTTCTCTTGA	5264
CCTTCCAACC CCCTTCCCAT AG AT GAC TTT ACT GAC TAT AAG AAG GGG GAA Asp Asp Phe Thr Asp Tyr Lys Lys Gly Glu 230 235	5315

CCC AAG CAC GGG AAG AGG AAA CGG GGG CGT CCC AGA AAG CTG AGC AAG Pro Lys His Gly Lys Arg Lys Arg Gly Arg Pro Arg Lys Leu Ser Lys 240 245 250 255	5363
GAA TAC TGG GAC TGT CTG GAG GGC AAG AAG AGC AAG CAC G Glu Tyr Trp Asp Cys Leu Glu Gly Lys Lys Ser Lys His 260 265	5403
GTAAGCTCTA AGGGCTGCCA GGCCTGTGGG CGGAGGGATA CTATTCCTTC AGCTTCCACT	5463
GGCCTCTCAC AGCCGCTGGA ACTCATTGCA TTGACGGGGC TCCATGGCAT TTGTTACTGC	5523
CTCTTTACAG AGGCCTGCTT GGA CTTAGAG AAGAAGGGAA CTGAGGTCCT AGGAGAGGCC	5583
ATGGAGAGAG TCCAGCCTTC CCACATTCTT CCTCTTTAAC TATCCCTGTA CTTGGCCCCCT	5643
GTCTTGCCCTG ATGGAAC TTT CCGATGGGGA GGAGGCAGCT GGTGGGTGCT CAGAGCCAGG	5703
CAGGCTGGGG TGGCTGCGAC TCCAGGCACA GCCTGCTGAA AGAGCCTCGG GCCTTGTGTG	5763
CTCCAGCTGC TCTGAACCAC CTGGAGGTCA CCCCAGAGGT CTTGCCTGCC GCCTCTTCCA	5823
TACACACAGA CCTGCACGTA TGCACTCCTG GCCTGCTTTC CTTCTTGTCC CGAGGGGACT	5883
GTTAATTCCG GGAAGCTGTT TCTTGGTCCC TCAGGCTATA GGCAGCTCTC TGACCCCATG	5943
TGTGCCAAGT TCTCACCACC ACTGGTCCCC ACTGAACCAT GAGCCCCCTC ACAAAGAAGC	6003
GTGTCTCTGT CGCTGTCCAT CTTAACCAGT TGTTTGATCC TTAAGTGGTG AGAGAATCGA	6063
GCGCTCTGTG CAGTCGGCCT AGCGCATTCG ATTTTGGGGC AGGAAAGGAA GCAGCCACTA	6123
TAGCAATCAC TAAGAGGACA TTTCATATAC TCCCATATGC CTTGGCTCTT AGCCTCGTTG	6183
GGATAGGAGA GGCCAGGTCG CCTAGAGGAG AGGGGCACCC CAGACTGATA ACTGAGGAAA	6243
TCTTCCCTTG TAG CC CCC AGA GGT ACT CAC CTG TGG GAG TTT ATC CGA Ala Pro Arg Gly Thr His Leu Trp Glu Phe Ile Arg 270 275 280	6291
GAC ATC CTA ATC CAC CCC GAG CTC AAC GAA GGC CTC ATG AAG TGG GAG Asp Ile Leu Ile His Pro Glu Leu Asn Glu Gly Leu Met Lys Trp Glu 285 290 295	6339
AAC CGG CAC GAG GGT GTG TTC AAG TTT CTT CGC TCA GAG GCC GTG GCC Asn Arg His Glu Gly Val Phe Lys Phe Leu Arg Ser Glu Ala Val Ala 300 305 310	6387
CAA CTC TGG GGC CAG AAG AAG AAG AAC AGC AAC ATG ACC TAT GAG AAG Gln Leu Trp Gly Gln Lys Lys Lys Asn Ser Asn Met Thr Tyr Glu Lys 315 320 325	6435
CTG AGC CGA GCC ATG AG GTGAGTGTGA GCGTCAGGGA CCTCTGCTTG Leu Ser Arg Ala Met Arg 330	6482
GGCTCTACTG GCTTCCGCTA GGTTTCACGA GACAGGCCTG AGGCCCGTAT GGAGAGGACA	6542
AGGACAGTGT TGTGGCCCTG TGTAAGTTGGT TACGTGCAGC ATGAAGAAAG CGCTGGGCAG	6602
AGATCGTGAG CACACTTAGC TTTAGCTAAC ATTTCTGTGT TTCCTGCAGA CTTGTTCTAA	6662
GAAAGACACT TGAGAGAGAG AAAGAATAGA AATTGACAGC TCAGCTCCCT TGTCTCTGGG	6722
CCACAAAGGT GAAGTAGCTC AGCATTGCTA AAGTCCCCTC TCCCTCAGTT CACGGGCCTT	6782
TATGAAAAGC CCCAGGACAT AGCCAGAAGG CACAGAGAAG TAAATGTAGA AGCAGGTGCT	6842

CTGGCCATAA TTACAGATCA CCGCGGCCAC AACAGGTGAG GAGAGGGAAC ACTCAGGCAG	6902
AGAGGGCCAG CTCAGCACAC TGGGGCTGGG AACCAATGCG AACCTCAGTC CATAGCATGC	6962
CTCTTGCTTA CACCTCTGAC CACCTCCTTC CCACGCAG G TAT TAC TAC AAA CGG	7016
	Tyr Tyr Tyr Lys Arg 335
GAG ATC CTG GAA CGG GTG GAT GGC CGA CGG CTC GTC TAC AAG TTT GGC	7064
Glu Ile Leu Glu Arg Val Asp Gly Arg Arg Leu Val Tyr Lys Phe Gly	
340 345 350 355	
AAG AAC TCT AGT GGC TGG AAG GAA GAA GAG GTT GGA GAG AGT CGG AAT	7112
Lys Asn Ser Ser Gly Trp Lys Glu Glu Glu Val Gly Glu Ser Arg Asn	
360 365 370	
TAAGGATCGG GGCTGGACCC AGGACCTGAC TCAGGCATGA ACTCCAGAAC TGAAGCCTTC	7172
CTGGAAGGAC AGGCAGGCCT GACGGCCCCC TTAACATGGA TGTGTTCCCT GTGTTGCTGT	7232
AGAGAGGAAG AACCTGTTGG GCGTGCCCTC TGCAGTCTCC TCAAGTGCAG CCTTTGGCCT	7292
CTCTCCTCGC CCTCTTGGA TTACAAGCCC CGGGTTTGAA CCAACTTGTT CGATAACTCT	7352
TCCAGCTGTG ATTCCAGTTC CCTCCCGTCC CAACATGGAC TGCAAATGAG ACCCACCTGC	7412
AGATGCCTGG CCTCAGCCAA GGAGGCTGGG GAGACTGTGG CAGGAGACTG CAGGGACGGA	7472
GGGGACAGGG TTGTGTCCTC GGTACTTCCT GGACTGCCTT CCACCTCTTT GCTCAGTACT	7532
CAGGCTCCAC AGACGGGGGT CGGATCATCC CTAATTTATG TGCTATAAAT ATTCCAGGTG	7592
TATATAGAGA GCTATTTTTT CTAAAGCATT TCCCCTCCCT GCTCTTCTCC ACTGAGTGCT	7652
GGTGGCCAGA CTGATTTTTT TTTTAGCCCC CCTAACTGGA CCAGCGAGAA GTAGGGTGAT	7712
TCCAGGACCC CCTCTTCCCC CAGAGGGGTC TCCTGGATCC	7752

(2) INFORMATION FOR SEQ ID NO:16:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 371 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:

Met	Ala	Ala	Thr	Cys	Glu	Ile	Ser	Asn	Val	Phe	Ser	Asn	Tyr	Phe	Asn
1				5					10					15	
Ala	Met	Tyr	Ser	Ser	Glu	Asp	Pro	Thr	Leu	Ala	Pro	Ala	Pro	Pro	Thr
			20					25					30		
Thr	Phe	Gly	Thr	Glu	Asp	Leu	Val	Leu	Thr	Leu	Asn	Asn	Gln	Gln	Met
		35				40					45				
Thr	Leu	Glu	Gly	Pro	Glu	Lys	Ala	Ser	Trp	Thr	Ser	Glu	Arg	Pro	Gln
	50					55				60					
Phe	Trp	Ser	Lys	Thr	Gln	Val	Leu	Glu	Trp	Ile	Ser	Tyr	Gln	Val	Glu
65					70				75					80	
Lys	Asn	Lys	Tyr	Asp	Ala	Ser	Ser	Ile	Asp	Phe	Ser	Arg	Cys	Asn	Met
				85					90					95	

Asp Gly Ala Thr Leu Cys Ser Cys Ala Leu Glu Glu Leu Arg Leu Val  
 100 105 110  
 Phe Gly Pro Leu Gly Asp Gln Leu His Ala Gln Leu Arg Asp Leu Thr  
 115 120 125  
 Ser Asn Ser Ser Asp Glu Leu Ser Trp Ile Ile Glu Leu Leu Glu Lys  
 130 135 140  
 Asp Gly Met Ser Phe Gln Glu Ser Leu Gly Asp Leu Gly Pro Phe Asp  
 145 150 155 160  
 Gln Gly Ser Pro Phe Ala Gln Glu Leu Leu Asp Asp Gly Arg Gln Ala  
 165 170 175  
 Ser Pro Tyr Tyr Cys Ser Thr Tyr Gly Pro Gly Ala Pro Ser Pro Gly  
 180 185 190  
 Ser Ser Asp Val Ser Thr Ala Arg Thr Ala Thr Pro Gln Ser Ser His  
 195 200 205  
 Ala Ser Asp Ser Gly Gly Ser Asp Val Asp Leu Asp Leu Thr Glu Ser  
 210 215 220  
 Lys Val Phe Pro Arg Asp Asp Phe Thr Asp Tyr Lys Lys Gly Glu Pro  
 225 230 235 240  
 Lys His Gly Lys Arg Lys Arg Gly Arg Pro Arg Lys Leu Ser Lys Glu  
 245 250 255  
 Tyr Trp Asp Cys Leu Glu Gly Lys Lys Ser Lys His Ala Pro Arg Gly  
 260 265 270  
 Thr His Leu Trp Glu Phe Ile Arg Asp Ile Leu Ile His Pro Glu Leu  
 275 280 285  
 Asn Glu Gly Leu Met Lys Trp Glu Asn Arg His Glu Gly Val Phe Lys  
 290 295 300  
 Phe Leu Arg Ser Glu Ala Val Ala Gln Leu Trp Gly Gln Lys Lys Lys  
 305 310 315 320  
 Asn Ser Asn Met Thr Tyr Glu Lys Leu Ser Arg Ala Met Arg Tyr Tyr  
 325 330 335  
 Tyr Lys Arg Glu Ile Leu Glu Arg Val Asp Gly Arg Arg Leu Val Tyr  
 340 345 350  
 Lys Phe Gly Lys Asn Ser Ser Gly Trp Lys Glu Glu Glu Val Gly Glu  
 355 360 365  
 Ser Arg Asn  
 370

(2) INFORMATION FOR SEQ ID NO:17:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 40 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide



(ix) FEATURE:  
 (A) NAME/KEY: Peptide  
 (B) LOCATION: 1..40  
 (D) OTHER INFORMATION: /note= "human ESX A-region/Pointed domain (amino acids 64-103 of SEQ ID NO:2)"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:

```

Pro Gln Phe Trp Ser Lys Thr Gln Val Leu Asp Trp Ile Ser Tyr Gln
1          5          10          15
Val Glu Lys Asn Lys Tyr Asp Ala Ser Ala Ile Asp Phe Ser Arg Cys
          20          25          30
Asp Met Asp Gly Ala Thr Leu Cys
          35          40
  
```

(2) INFORMATION FOR SEQ ID NO:18:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 38 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS:  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(ix) FEATURE:  
 (A) NAME/KEY: Peptide  
 (B) LOCATION: 1..38  
 (D) OTHER INFORMATION: /note= "human ETS-1 A-region/Pointed domain (amino acids 69-106)"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:

```

Pro Arg Gln Trp Thr Glu Thr His Val Arg Asp Trp Val Met Trp Ala
1          5          10          15
Val Asn Glu Phe Ser Leu Lys Gly Val Asp Phe Gln Lys Phe Cys Met
          20          25          30
Asn Gly Ala Ala Leu Cys
          35
  
```

(2) INFORMATION FOR SEQ ID NO:19:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 51 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS:  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(ix) FEATURE:  
 (A) NAME/KEY: Peptide  
 (B) LOCATION: 1..51  
 (D) OTHER INFORMATION: /note= "human ESX serine-rich box (amino acids 188-238 of SEQ ID NO:2)"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:

Ala Pro Ser Pro Gly Ser Ser Asp Val Ser Thr Ala Gly Thr Gly Ala  
1 5 10 15  
Ser Arg Ser Ser His Ser Ser Asp Ser Gly Gly Ser Asp Val Asp Leu  
20 25 30  
Asp Pro Thr Asp Gly Lys Leu Phe Pro Ser Asp Gly Phe Arg Asp Cys  
35 40 45  
Lys Lys Gly  
50

(2) INFORMATION FOR SEQ ID NO:20:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 51 amino acids  
(B) TYPE: amino acid  
(C) STRANDEDNESS:  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(ix) FEATURE:

(A) NAME/KEY: Peptide  
(B) LOCATION: 1..51  
(D) OTHER INFORMATION: /note= "SOX4 serine box (amino acids 370-420)"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:20:

Ala Pro Ser Ser Ala Pro Ser His Ala Ser Ser Ser Ala Ser Ser His  
1 5 10 15  
Ser Ser Ser Ser Ser Ser Ser Gly Ser Ser Ser Ser Asp Asp Glu Phe  
20 25 30  
Glu Asp Asp Leu Leu Asp Leu Asn Pro Ser Ser Asn Phe Glu Ser Met  
35 40 45  
Ser Leu Gly  
50

(2) INFORMATION FOR SEQ ID NO:21:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 16 amino acids  
(B) TYPE: amino acid  
(C) STRANDEDNESS:  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(ix) FEATURE:

(A) NAME/KEY: Peptide  
(B) LOCATION: 1..16  
(D) OTHER INFORMATION: /note= "portion of human ESX serine box showing clustering of serine residues opposite a hydrophobic face in a helical wheel model"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:21:

Ser	Pro	Gly	Ser	Ser	Asp	Val	Ser	Thr	Ala	Gly	Thr	Gly	Ala	Ser	Arg
1				5					10					15	

(2) INFORMATION FOR SEQ ID NO:22:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 81 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS:

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(ix) FEATURE:

(A) NAME/KEY: Peptide

(B) LOCATION: 1..81

(D) OTHER INFORMATION: /note= "human ESX Ets DNA binding domain (amino acids 274-354 of SEQ ID NO:2)"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:22:

His	Leu	Trp	Glu	Phe	Ile	Arg	Asp	Ile	Leu	Ile	His	Pro	Glu	Leu	Asn
1				5					10					15	

Glu	Gly	Leu	Met	Lys	Trp	Glu	Asn	Arg	His	Glu	Gly	Val	Phe	Lys	Phe
		20					25						30		

Leu	Arg	Ser	Glu	Ala	Val	Ala	Gln	Leu	Trp	Gly	Gln	Lys	Lys	Lys	Asn
		35				40						45			

Ser	Asn	Met	Thr	Tyr	Glu	Lys	Leu	Ser	Arg	Ala	Met	Arg	Tyr	Tyr	Tyr
	50					55					60				

Lys	Arg	Glu	Ile	Leu	Glu	Arg	Val	Asp	Gly	Arg	Arg	Leu	Val	Tyr	Lys
65					70				75						80

Phe

(2) INFORMATION FOR SEQ ID NO:23:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 81 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS:

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(ix) FEATURE:

(A) NAME/KEY: Peptide

(B) LOCATION: 1..81

(D) OTHER INFORMATION: /note= "Elf-1 Ets DNA binding domain (amino acids 209-289)"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:23:

Tyr Leu Trp Glu Phe Leu Leu Ala Leu Leu Gln Asp Lys Ala Thr Cys  
1 5 10 15  
Pro Lys Tyr Ile Lys Trp Thr Gln Arg Glu Lys Gly Ile Phe Lys Leu  
20 25 30  
Val Asp Ser Lys Ala Val Ser Arg Leu Trp Gly Lys His Lys Asn Lys  
35 40 45  
Pro Asp Met Asn Tyr Glu Thr Met Gly Arg Ala Leu Arg Tyr Tyr Tyr  
50 55 60  
Gln Arg Gly Ile Leu Ala Lys Val Glu Gly Gln Arg Leu Val Tyr Gln  
65 70 75 80  
Phe

(2) INFORMATION FOR SEQ ID NO:24:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 6 amino acids  
(B) TYPE: amino acid  
(C) STRANDEDNESS:  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:24:

Leu Trp Gln Phe Leu Leu  
1 5

(2) INFORMATION FOR SEQ ID NO:25:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 4 amino acids  
(B) TYPE: amino acid  
(C) STRANDEDNESS:  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:25:

Lys Leu Ser Arg  
1

(2) INFORMATION FOR SEQ ID NO:26:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 5 amino acids  
(B) TYPE: amino acid  
(C) STRANDEDNESS:  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:26:

Leu Arg Tyr Tyr Tyr  
1 5

(2) INFORMATION FOR SEQ ID NO:27:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 4 amino acids  
(B) TYPE: amino acid  
(C) STRANDEDNESS:  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:27:

Leu Trp Glu Phe  
1

(2) INFORMATION FOR SEQ ID NO:28:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 4 amino acids  
(B) TYPE: amino acid  
(C) STRANDEDNESS:  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:28:

Arg Tyr Tyr Tyr  
1

(2) INFORMATION FOR SEQ ID NO:29:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 4 amino acids  
(B) TYPE: amino acid  
(C) STRANDEDNESS:  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:29:

Arg Leu Val Tyr  
1

(2) INFORMATION FOR SEQ ID NO:30:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 397 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:30:

TCAGCCCTGG CCAGGCCCCC AGGAAGAATT TCCAGGGCCA GAGGGCAGCC TAAGGCACAG	60
ATGCCCACCC CTGCAATGTT CCCGCCACAT GCCCAGTTCA GTACCCAGGG CCCAACCCCA	120
GAGGGTGCGG AATGACAGAT TCTGACAATC ATTAAACCAG CCAGGCCTGA TTTCCCAGCA	180
CCGCCCCTTA GGATATGGGC CAAGTGGCAC GGAATATGCA AATCACATGG GACAGGGAGC	240
CCAGTCTGAA GGCCAGGAAA TCCCCAGCAT CCAATGAGCC ACCAGCTCAG GTTACAACCG	300
GGGACGTACG CCGAAGACCT GGAGGGGAGG AGCTCCTGCT TTGCTCTATT TAGAGCGGGT	360
GGGGGCAGCG CCCTGGCCAC ACTCATCACT GCTACCT	397

(2) INFORMATION FOR SEQ ID NO:31:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 400 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:31:

CCATCTCTGG CCTGGCCCCCT GGGAGGAATT TCCTGGGCCA GAGGGCAGCC GAAAGCACAG	60
ATGCCCACCC CAGCAACGTT CCCGCCACCT GCCCAGGCCA GTGCCCCGTG CCCAACCCCA	120
GAGGGTGCGG GATGACAGAC TCTGACAATC ATTAAACCAG CCGGGCCTGA TTTCCCAGCA	180
CTGCCTGCTA AGATCCGGGC CAAGTGGCAC TGAATATGCA AATCACATGG GGCCAGGAGC	240
CCAGTCTAAA GGCCAGGAAA TCCCCTCCAT CCAATGAGAC ACCAGCTCAG GTTACTGCAG	300
GGGACACACT ATAAAGCCCT GAGCTCAGGG AGGAGCTCCC TCCAGGCTCT ATTTAGAGCC	360
GGGTAGGGGA GCGCAGCGGC CAGATACCTC AGCGCTACCT	400

(2) INFORMATION FOR SEQ ID NO:32:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 31 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:

- (A) NAME/KEY: -
- (B) LOCATION: 1..31
- (D) OTHER INFORMATION: /note= "TA5 oligonucleotide containing Ets responsive element from HER2/neu promoter"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:32:

GGAGGAGGGC TGCTTGAGGA AGTATAAGAA T	31
------------------------------------	----

(2) INFORMATION FOR SEQ ID NO:33:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 31 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

- (ix) FEATURE:  
 (A) NAME/KEY: -  
 (B) LOCATION: 1..31  
 (D) OTHER INFORMATION: /note= "m1 mutant TA5 sequence"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:33:

GGAGGAGGTA TGCTTGAGGA AGTATAAGAA T

31

(2) INFORMATION FOR SEQ ID NO:34:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 31 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

- (ix) FEATURE:  
 (A) NAME/KEY: -  
 (B) LOCATION: 1..31  
 (D) OTHER INFORMATION: /note= "m2 mutant TA5 sequence"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:34:

GGAGGAGGGC TGCTTGCGGA AGTATAAGAA T

31

(2) INFORMATION FOR SEQ ID NO:35:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 31 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

- (ix) FEATURE:  
 (A) NAME/KEY: -  
 (B) LOCATION: 1..31  
 (D) OTHER INFORMATION: /note= "m3 mutant TA5 sequence"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:35:

GGAGGAGGGC TGCTTGAGAG AGTATAAGAA T

31

(2) INFORMATION FOR SEQ ID NO:36:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 31 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(ix) FEATURE:

- (A) NAME/KEY: -  
 (B) LOCATION: 1..31  
 (D) OTHER INFORMATION: /note= "m4 mutant TA5 sequence"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:36:

GGAGGAGGGC TGCTTGACCA AGTATAAGAA T

31

(2) INFORMATION FOR SEQ ID NO:37:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 31 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(ix) FEATURE:

- (A) NAME/KEY: -  
 (B) LOCATION: 1..31  
 (D) OTHER INFORMATION: /note= "m5 mutant TA5 sequence"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:37:

GGAGGAGGGC TGCTTGAGGA AGCATAAGAA T

31

(2) INFORMATION FOR SEQ ID NO:38:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 15 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:38:

ATACTTCCTC AAGCA

15